图书情报工作
第 64 卷 第 5 期 2020 年 3 月

基于修正点互信息的特征级情感词极性自动研判*

■ 聂卉 首欢容
中山大学资讯管理学院 广州 510275

摘 要:[目的/意义] 基于语料的情感词发现依语句上下文推断情感词极性,能显著提升情感分析的准确率,在面向领域的特征级情感分析任务中有重要应用价值。[方法/过程] 对特征级情感极性研判问题展开探讨,提出基于点互信息的"特征-情感"对情感极性自动判别算法,算法借助大规模领域语料,根据观点表达"特征-情感"对与情感语义明确的种子词的共现关系,同时引入依存句法分析语句间的情感转折,通过修正经典的点互信息算法,对上下文约束下的用户观点表达进行褒贬预测。[结果/结论] 实验证明,修正算法的性能显著优于词典匹配算法和经典的点互信息情感识别算法,不仅能够推断词典中未纳入的观点表达的情感指向,而且能较准确地推断"语境"中的情感词极性。在餐饮评论和数码产品评论两个评测语料集上,修正算法的 F1 宏平均指标分别达到 0.827 和 0.878。该算法以领域相关的大规模语料为支撑,基于概率统计和句法分析,因数据获取便利,算法效率高,移植性好,具有普适性,尤其适用于面向领域的情感分析任务。

关键词:情感分析 点互信息算法 领域情感词 上下文
分类号:P393.092 O212.1
DOI:10.13266/j.issn.0252-3116.2020.05.012

1 引言

Web 技术和电子商务的飞速发展正极大改变着人们的生活,人们不仅习惯于在线购物,而且常常在平台发表购物体验,网络上由此产生了大量的在线评论。用户的在线产品评论在商务领域被称为产品的网络口碑,因富含重要商业价值,受到业界和学术界的关注。

从在线评论中提取有价值信息是"评论挖掘"的重要内容。评论挖掘针对蕴含观点的主观性文本,基本任务是判断用户对产品各特征性能所发表评价的情感倾向——正向或负向。这一情感分析任务面向产品特征,情感词极性往往与评价对象有关。如,在手机评论中,对于形容词"高","性价比-高"为正向,"能耗-高"为负向;在股票评论中,"通胀率-高"为负向,"利率-高"则为正向。可见,某些观点词的情感指向会随关联对象发生变化,出现所谓的"情感歧义"。

对于特征级情感分析,具有"情感歧义"属性的情感词的出现会使得基于词典的情感极性研判产生偏差。因为,在情感词典中,词项的极性均被固化;但实际应用中,诸多具有情感色彩的词项,其极性会随"语境"发生变化("语境"在本文主要指情感词所在语句的上下文语义约束,如修饰对象),如"大,小,高,低"等,对于这类词,通用词典通常不收纳,或将其再细分,如"大连理工大学中文情感词汇本体库"[1]的情感词分为"褒,贬,褒贬两性,中性"4 类。褒贬词有明确的情感极性,中性词和褒贬两性词的情感极性则与"语境"相关。对于特征级情感分析,能够准确判断随"语境"变化的词的情感极性对确保精准度、提升分析质量有重要意义。许多研究将其归为面向领域的情感词典构建问题,且普遍认为,情感词项有领域依赖性,基于领域语料发掘情感词并对其极性进行研判是提升情感分析精度的有效手段。研究者们已注意到词项的情感歧义问题,但领域情感词典构建针对的仍是孤立情感词,歧义问题并未得到根本解决。

本研究关注情感词极性的歧义问题,认为要妥善解决褒贬两性或中性词的情感指向不定问题,需将情感词与关联对象(特征)作为整体考虑,同时将分析对象所在语句的上下文约束关系纳入分析模型。因此,

* 本文系国家社会科学基金项目"面向用户感知效用的在线评论的质量与控制研究"(项目编号:15BTQ067)研究成果之一。
作者简介:聂卉(ORCID:0000-0001-8567-3084),副教授,博士,E-mail:issnh@mail.sysu.edu.cn;首欢容(ORCID:0000-0003-0586-218X),硕士研究生。
收稿日期:2019-04-23 修回日期:2019-10-21 本文起止页码:114-123 本文责任编辑:易飞

本研究强调“特征－情感”对的情感极性研判，研究利用大规模领域语料，并借鉴点互信息算法，通过引入依存句法捕获情感表达的上下文约束来对“特征－情感”的极性进行推断。算法选择非监督策略，以减少人工介入；而算法基于领域语料，以使其具有更好的灵活性。

2 相关研究

2.1 面向评论的特征级情感分析

面向评论的特征级情感分析为细粒度情感分析。如一则手机评论“这款手机的散热太差了，玩一会儿就烫手，像素也就一般吧，但是整体运行速度还不错，在这个价位的手机中还算值得推荐”。评论中，用户对该手机的性能进行了多方点评，特征级情感分析则需从中提取“散热－太差”“像素－一般”“运行速度－不错”等一组用户观点，并准确判断其情感指向。

早期研究中，M. Q. Hu 等[2] 和 B. Liu 等[3] 采用关联规则挖掘产品特征词，并通过 WordNet 扩充形容词确定用户观点(情感词)，进而预测用户对产品性能的情感倾向；A. M. Popescu 等[4] 通过引入语法规则，更好地实现了“特征－情感”对的识别。对于特征级情感分析，除需有效提取“特征－情感”对，另一关键问题是情感值计算，一般情况下，词粒度的情感极性判断和计算都需借助情感词典。中文领域，HowNet 和大连理工中文情感词汇本体[1] 均是被广泛采用的通用情感词典。但通用词典多为人工生成，词项覆盖率有限，实际应用中，需要根据应用领域拓展词表，构建领域情感词典。代表性研究中，P. D. Turney 等[5] 提出利用情感种子词计算词项间的互信息用于推测拓展词的褒贬倾向，点互信息(Point Mutual Information, PMI)法在情感分析领域被广泛采纳，情感词典在情感分析领域发挥了重要作用。

但随着研究的深入，有学者[6-8] 注意到不同释义下情感词可能有不同的情感指向，如“长”在描述电池续航时间和聚焦时间上的情感极性是相反的[9]，“高”在针对“性价比”和“物价”的评价中，情感极性也出现了歧义；学者 X. Ding 等[6] 特别指出对于特征级情感分析，情感词极性要与所修饰的特征词形成对应，才能确保分析精度。随后，许多研究开始关注面向特征的情感词极性的准确判定问题。Y. F. Wu 等[8] 提出基于常识的情感词消歧方法，针对 14 个中文高频特征依赖形容词，如“大小多少高低”，依据对特征词和修饰形容词的副词(“有点”“那么”)的判断，使用固定模式判断其情感指向。A. Balahur 等[10] 基于 3 种策略的多数投票结果来对“特征－情感”对的极性进行研判，3 种策略分别是基于上下文的有监督学习、基于网络查询的最高点击以及基于规则的方法；H. H. Lek 和 D. C. C. Poo[11] 针对“特征－情感”对，首先进行特征词聚类，再利用 WordNet 对情感词和特征词的近义词进行合并，最终采用分类法对情感词指派极性；Y. Q. Xia 等[12] 则对情感词所在语句的内部要素进行了细致分析，根据句中的修饰副词及明确极性的情感词，运用贝叶斯分类法对有特征依赖性的情感词的极性进行判断。

另一方面，特征级的情感分析还强调情感计算的精细化，提出用户观点的量化不仅要准确辨明情感极性，还应考虑其修饰成分。例如“尤其贵”“不太友好”“还算满意”等。这些观点表述中，修饰成分加强、减弱甚至反置了情感词极性，若要提升分析的精准度，修饰副词的程度影响不可忽略。如 C. Whitelaw 等[13] 结合 WordNet 用半自动方法的同时构建了情感词词典及修饰词词典，根据修饰词的属性值来计算观点表达的情感值；S. L. Huang 等[14] 依据词性匹配模式定位“特征－情感”对，并通过引入副词及否定词来确定情感词极性，量化用户观点；史伟等[15] 深入探讨了副词的作用，采用加权原则，达到细化情感分值的目的；笔者[16] 也曾利用 Stanford Parser 语法工具获得语句中副词与情感词的修饰关系，继而通过设定修饰词与情感词的组配模式来确定观点表达的极性，实现细粒度的情感分析。

由此可见，特征级情感分析强调的是对用户褒贬观点的准确研判。用户观点比单纯的情感词具有更丰富的表达形式与内涵。特征级情感分析不仅要对“特征－情感”对统一考虑，还需对情感表述所在语句的语法结构和语义关系进行分析。

2.2 情感词极性研判与情感词典构建

本研究主要对“特征－情感”对的极性判断问题进行探究，主要任务是辨识情感词，并根据修饰对象判断情感词指向。这一工作亦是构建情感词典的核心任务。

2.2.1 基于词典的情感词发现与极性判断

词典拓展是自动生成情感词典的重要方法，该方法利用词典定义的词间关联，基本思路是：先确定一组极性明确的情感种子词(褒或贬)，再在词典中搜索种子词的同/反义词，谓之拓展词，根据拓展词与种子词的同/反义关系赋予拓展词极性，纳入情感种子集，这一过程循环迭代，直至情感词集收敛。M. Q. Hu 和 B.

Liu 采用这一方法基于 WordNet 生成了广被采纳的 Opinion Lexicon[2]；A. Esuli 和 F. Sebastian [17]认为，情感指向一致的词项极可能有相似的注释信息，对 WordNet 的注释进行分类能够帮助判断词项极性；J. Kamps 等[18]先建立基于 WordNet 的同义词网络，然后借助网络结构推断候选词的情感倾向，在同义词网络中，候选词与“good”或“bad”间的距离决定其褒贬性质；A. Hassan 和 D. Radev[19]则采用了先进的马尔科夫随机游走算法，基于 WordNet 近义和上位关系生成词图，通过计算候选词跳转到褒贬词集的平均步数，指派其情感指向。整体看来，这类研究通常将候选词作为独立单元，通过计算其与情感语义明确的种子词的关联度推断其褒贬性质。基于词典研判词项极性往往受词典定义制约，词项极性随修饰对象及语句上下文会发生变化的问题不能得以解决，词典方法对情感语义明确的词项的识别效果较好，对褒贬兼容或中性词的情感推断则可能出现偏差。

2.2.2　基于领域语料的情感词挖掘与极性判断

基于词典拓展获得的情感词极性受制于词典，而诸多研究面向的是特定领域的情感分析任务，利用领域语料进行情感词挖掘更具实用价值。这类方法从领域语料中发掘情感词，通过构建领域情感词典提升情感分析精度。如 N. Oliveira 等[20]标注了一个基于股票推特的语料库，借助语料库计算候选词与“牛/熊市”的关联度，从而发现股票领域特有的情感词，该方法需事先对语料进行标注；S. Deng 等[21]则提出通过无标注数据构建领域词典，他们先对情感候选词做细致筛选，再运用点互信息推断候选词的极性，该研究因引入了大规模无标注领域语料，建模效率高，提升了词典在执行领域情感分析任务时的实效；郗亚辉[22]的领域情感词典构建分两个阶段，首先利用词的点互信息和上下文约束关系建立词矩阵，在矩阵上运用标签传播算法拓展种子构造基本情感词典，再根据语料中出现的情感冲突频率来判断领域关联情感词的极性；K. Labille 等[23]基于概率论和信息论，借助评论语料构建了 15 个领域的情感词典，实际应用中，表现优良。致力于领域情感词典构建的研究均认为领域词典具有针对性，能解决情感歧义问题，比通用情感词典更有效。深度学习方法也被用于领域情感词典构建：Y. Li 等[24]提出在词向量训练过程中加入更多的先验知识优化词嵌入矩阵，使其能够适应特定领域的情感分析任务；林江豪[25]等采用基于大规模语料构建 W2V(Word2Vector)的预训练词向量对词项进行深度表示，结合情感词典，得到词的情感向量表示，赋予词项更丰富的情感语义。

总体看来，基于语料的方法对具有领域依赖性的情感词研判比较适用，非标注语料在实际应用也更具可行性。但单纯生成领域词典，还是不能改变某些词项(褒贬双性或中性词)的“语境”依赖性。本研究所指“语境”特指情感词所在语句的上下文约束，尤其是情感词修饰的特征词。对于这类情感词极性研判问题，本研究认为，只有将情感词所在语句的上下文及修饰对象纳入模型，才可能对其情感指向做出准确推断。为此，与构建领域词典的做法不同，研究将“特征－情感”对作为情感识别对象，就整个“观点表达”进行极性推断，并对点互信息算法进行修正，将基于依存句法分析获得的“情感转折”状态纳入计算模型。本研究提出的方案因兼顾了修饰对象及语句的上下文约束，能处理情感歧义问题，并使分析精度得以提升；另则，本研究采用基于语料的情感词挖掘方法，不受词典约束，在领域语料充沛的前提下，该方法能够非常方便地迁移至不同场景。大数据环境下，获取无标注领域语料成本低且十分便利，因而本方法具有较好的普适性。

3　基于点互信息的“特征－情感”对极性研判

3.1　互信息及基于点互信息的情感词极性判断

互信息是信息论中的概念，在文本挖掘领域被用于计算词汇间的相关度。点互信息计算基于共现概率，语料充足的前提下，词项共现率越高，相关性越大。若以 w1 与 w2 表示两词项，其互信息计算公式如下：

$$PMI(w1,w2)=\log_2\left(\frac{P(w1,w2)}{P(w1)P(w2)}\right) \tag{1}$$

其中，P(w1)与 P(w2)为词项独立出现的概率，P(w1,w2)为二者共现概率。P. D. Turney 和 M. L. Littlema[5]将互信息引入情感分析，称为情感点互信息计算(Semantic Orientation Pointwise Mutual Information，SO-PMI)。其依据是：情感倾向相同的词项常频繁出现，如“开心－喜悦”“痛苦－悲伤”，情感倾向相反的词则不然。这使得“开心”与“喜悦”的互信息较高；与“痛苦”的互信息则较低。因此，对于一个待测情感词项 w，将其和褒义种子词集 Pos_set 的互信息与其和贬义种子词集 Neg_set 的互信息进行比较，可推测 w 的情感指向，有公式(2)：

$$SO_PMI(w)=\sum_{w_P\in Pos_set}PMI(w,w_P)-\sum_{w_N\in Neg_set}PMI(w,w_N) \tag{2}$$

若 SO-PMI(w)＞0，w 的情感极性偏正；SO-PMI

聂卉，首欢容.基于修正点互信息的特征级情感词极性自动研判[J].图书情报工作,2020,64(5):114－123.

(w) < 0；w 的情感极性偏负。可见，经典的点互信息情感计算与情感种子集 Pos_set 和 Neg_set 关联密切，且情感依存关系设定相对简单。但真实语言环境中，情况往往比较复杂。极性相反的词也时常同时出现，如"痛并快乐""乐极生悲"等。对于这类情况，运用单纯的 SO_PMI 计算方法肯定会产生偏差。

另则，本研究以"特征－情感"为极性研判对象，非纯粹情感词，基于共现关系的情感一致性会因句式结构产生变化。如"这家餐厅的价钱合理，但是服务太差了。"该评论从两方面评价一家餐厅，"价钱－合理"为褒，"服务－差"为贬。褒贬情感状态共现，若以情感极性一致为准则计算 PMI，则无论是通过"服务－太差"去推断"价钱－合理"的极性，还是基于"价钱－合理"去推测"服务－太差"的极性，都可能出现错判。分析语句发现，虽然共现，但待测"特征－情感"的情感指向因语句中存在转折词"但"发生了逆转。类似反向情感共现的情况，语言中十分普遍，各种转折关系的介入导致主体词尽管一致，但语义差别甚远，如"价格合理，但是服务太差。""价格合理，服务也不差。""价格不合理，服务也很差。""价格不合理，但是服务不差。"，这 4 句话在语义和情感上存在明显区别，在只考虑共现的 PMI 计算中，4 句话里都有"价格－合理"和"服务－差"的出现，使得这 4 句话的 PMI 计算等价，但显然，如此处理将导致错判。本研究认为，对"特征－情感"的极性进行推断，在共现基础上，还应将语句表达中的情感反向及转折关系纳入分析中，利用句法依存分析可更准确地提取词项间的情感关系，减少错判几率，提升算法精度。

3.2 基于点互信息的"特征－情感"极性研判修正算法

本研究对 P. D. TURNEY [5] 提出的情感互信息计算方法进行了改进，针对"特征－情感"的情感倾向性判断，采用依存句法分析，通过引入反转/转折关系来修正"特征－情感"与种子情感词集的情感一致性，以减少识别偏误。研究先通过观察，对语料中情感转折现象进行分析，示例如表 1 所示：

表 1　否定/转折词对语句情感语义的影响(示例)

示例	转折词	否定词	是否转折	极性关系	极性判断
价格合理，但是服务太差	但是	－	是	情感极性不一致	服务－差：反向价格－合理：正向
价格合理，服务也不差	－	不	是	情感极性不一致	服务－差：反向价格－合理：正向
价格不合理，服务也很差	－	－	否	情感极性一致	服务－差：反向价格－不合理：反向
价格不合理，但是服务不差	但是	不	否	情感极性一致	服务－差：反向价格－不合理：反向

注：设"差"为负向情感种子，"价格－合理"，"价格－不合理"为待测特征情感组配

表 1 中，"但是"和"不"均表示发生了情感转折，共现窗口内，反转词为单数时，"特征－情感"的极性与种子情感词极性不一致，发生了情感反转。因此，借由"服务－差"与"价格－合理"的共现关系正确推断"价格－合理"的极性应根据反转情况修正"价格－合理"与褒贬种子集的互信息计算。算法改进的基本思路为：①将"特征－情感"引入 SO_PMI 计算，设共现窗口为邻接子句；②构建转折/否定词典，引入转折/否定词与待判"特征－情感"及种子词的依存关系，分析情感语义转折现象；③在计算"特征－情感"与褒贬种子词集的互信息时，基于共现及情感转折状态，修正"特征－情感"与种子词的情感关系。算法描述如下：

设 *Corpus* 为分析语料，采用特征情感抽取算法从中提取"特征－情感"，记为 $<f,o>$；$<f,o>$ 与正/负面情感的关联度分别表示为 $PMI_{pos}(<f,o>)$，$PMI_{neg}(<f,o>)$；*Pos_set* 和 *Neg_set* 分别对应褒贬情感种子集；设 w 为情感种子词，$w \in Neg_set \cup Pos_set$。

输入：$<f,o>$，*Corpus*，*Pos_set*，*Neg_set*

输出：$<f,o>$ 的情感极性

算法描述：

Step1：对 $<f,o> \in Corpus$ 初始化，即 $PMI_{pos}(<f,o>)=0$，$PMI_{neg}(<f,o>)=0$；

Step2：对语料中的原始语句进行切分，设定共现窗口为切分后的短句，短句中包含 $<f,o>$ 和某种子情感词 w；

Step3：引入否定/转折词典，采用依存句法分析短句，针对 w 判断情感转折是否发生；

Step4：对于 $w \in Pos_set$ 计算"PMI"$(<f,o>,w)$值；若短句中情感未发生转折，有

$PMI_{pos}(<f,o>) = PMI_{pos}(<f,o>) + PMI(<f,o>,w)$

若情感发生了转折，则

$PMI_{neg}(<f,o>) = PMI_{neg}(<f,o>) + PMI(<f,o>,w)$

Step5：对于 $w \in Neg_set$ 计算 $PMI(<f,o>,w)$值；若短句中情感未发生转折，有

$PMI_{neg}(<f,o>) = PMI_{neg}(<f,o>) + PMI(<f,o>,w)$

若情感发生了转折，则

$PMI_{pos}(<f,o>) = PMI_{pos}(<f,o>) + PMI(<f,o>,w)$

Step6：计算 $<f,o>$ 的点互信息，有

$SO_PMI(<f,o>) = PMI_{pos}(<f,o>) - PMI_{neg}(<f,o>)$

若 $SO_{PMI}(<f,o>)>0$，表明 $<f,o>$ 与正面情感更紧密，可能表达正面情感；反之可能表达的是负面情感。

需特别说明的是，算法第 2 步（共现窗口提取），切分短句采用的是哈工大语言技术平台提供的 SentenceSplit 分句接口，该功能根据分号、叹号、句号等分隔符对长句进行切分，因此短句中仍可能包含子句，子句间以逗号分隔。本研究将短句为窗口单位，在短句上进行依存句法分析。另则，算法第 3 步，借由依存关系判断情感转折。如短句“价格合理，但是服务很差。”这句话的依存分析结果中，“但是”与“差”具有直接依存关系，于是将“但是”这一转折的发生累计在“差”这一负向情感词上。因此，窗口内“价格 – 合理”对于“差”发生的转折次数为单数（1），情感语义发生反转，“价格 – 合理”对应正向情感。

利用上述算法可直接通过分析语料推断“特征 – 情感”的情感指向。“特征 – 情感”为统一体，受共现窗口内上下文约束的影响，引入依存句法关系可增强判断，理论上可提升识别准确率。

4 实验与分析

4.1 实验设计及流程

研究的实验流程见图 1，分 4 部分：数据采集与预处理、种子情感词典及转折/否定词典构建（资源建设）、基于改进互信息的“特征 – 情感”情感极性自动判定。

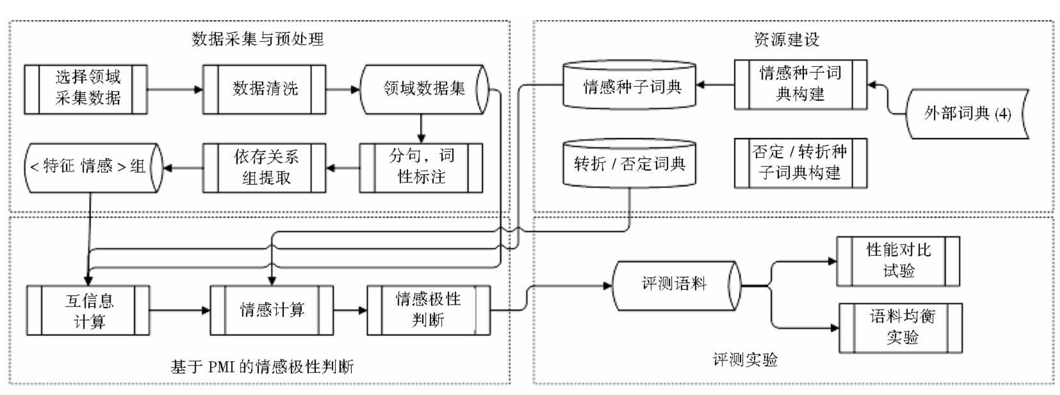


图 1 实验设计与流程

数据采集与预处理模块，因需检测算法的领域适用性，本研究选择餐饮与数码产品两个领域的评论数据为分析语料，主要针对评论内容，对原始数据进行清洗，采用自然语言预处理技术进行分句、分词、运用依存关系分析提取评论内容中的“特征 – 情感”。在资源建设模块，利用 4 个外部词典构建种子情感词集，人工建立转折/否定词词典；且以外包标注方式获得一份评测语料集。主体实验采用 3.2 中提出的优化算法对“特征 – 情感”进行情感倾向预测，并以经典的 SO_PMI 和基于词典的情感匹配法为参照，在评测语料集上进行实验，检验算法效能。

4.2 实验语料与资源建设

4.2.1 实验语料

研究选择数码产品和餐饮领域，采集真实评论语料进行分析。餐饮评论源自大众点评网，在网站根据排名选取广州地区粤菜商家的评论，编写爬虫爬取目标商家的用户评论内容与评价星级（评论时间跨度为 2004 年 7 月 7 日至 2017 年 10 月 31 日），共 400 126 条，其中差评（评价星级 < 3）23 663 条，好评（评价星级 > 3）376 463。数码产品评论源自京东商城，数据获取途径则是藉由下载他人分享的数据包（京东评论语料：https://download.csdn.net/download/qq_30365565/10657579）获得，下载数据为原始评论语料，进一步对语料进行了清洗、去重、过滤等一系列处理，京东数码产品评论的总量为 108 919 条差评，101 905 条好评。

4.2.2 词典

本研究整合了 4 个外部词典生成通用情感词典（https://download.csdn.net/download/weixin_42018090/10358043），分别是知网 HowNet 正/负面情感词语、NTUSD 台湾大学情感词典、清华大学李军中文褒贬义词典（TSING）和大连理工的情感词汇本体。对归并后的词典进行整理，剔除冗余和冲突情感词。研究所用的种子情感词集选自归并的情感通用词典，均为情感语义明确的情感词项，共 10 653 个正面词、15 888 个负面词。

情感反转判断需识别转折词和否定词，为此构建否定/转折情感词典。否定词比较明确，参考网上资源（https://download.csdn.net/download/fkyyly/10650404）本

研究共汇总了 38 个常用中文否定词,语义的转折往往发生在邻接语句间。一般情况,前面语句提出某事实或状态,后面的语句转而述说与前面语句相反的意思,后面语句通常表述说话人的真正意图。中文转折词归属连词,常用转折词数量不大,通过参阅网络资源,并观察从语料中随机抽取的语句,本研究以人工方式汇集了 15 个常用中文转折词,如表 2 所示:

表 2　否定/转折词典

否定词	转折词
不、没、否定、不能、反而、并非、相反、无法、不够、不是、不曾、未必、不要、未曾、全无、不太、未免、不足、未经、可不、并不、不意、不意味、不意味着、不足以、不怎么、反过来、并不一定、没有、不会、毫无、毫不、并不比、绝不、不可、没什么、难以、不算	可是、但是、虽是、虽说、尽管、固然、不过、只是、虽然、却、然而、就是、除了、但不、但

4.2.3　评测语料

为验测算法性能,本研究专门构建了一份评测语料,分别从京东电子产品评论集和大众点评餐饮业评论集中随机抽取评论 1 000 条和 1 600 条,以外包形式委托数据公司进行人工标注。每则评论要求从中抽取“特征－情感”对,同时注明“特征－情感”的情感指向(正向和反向),标注结果采用 <特征,情感,情感极性>三元组形式保存。标注完成后,数据公司派专人随机抽样 10% 进行验测,提交检测报告,准确率达 90% 以上为满足标注的精度要求。实际评测语料随机验测的精度为 94% 。此外,为了确保语料均衡,在随机抽取的前提下好/差评数量大致相当。标注示例如表 3 所示:

表 3　测试语料标注示例

评论_ID	评论内容	标注示例
6013521	使用了一天才来评价运行速度都几快,就是电池不耐用,照相效果一般般	<运行速度,快,1> <电池,不耐用,－1>
5954708	手机很养眼,分辨率不行,玩游戏卡屏,连接 WIFI 的时候信号不好	<手机,养眼,1> <分辨率,不行,－1> <玩游戏,卡屏,－1> <信号,不好,－1>
375645990	果然是网红店,很多人,款式很新颖,味道也不错,价格偏贵,装修有特色,值得一来	<人,多,1> <款式,新颖,1> <味道,不错,1> <价格,贵,－1>
375497455	毕德寮,粤语的意思:好的不得了。网红茶点,点都德的精致版,茶点中的:“爱玛仕”,环境味道都一流	<味道,一流,1> <环境,一流,1>

注:1 代表正向情感,－1 代表负向情感

4.3　实验与结果分析

4.3.1　评测方法

算法性能评测采用经典的精准度 P,回召率 R 和 F1。为对算法进行有效评价,本研究采用了多分类指标计算,即分别计算正/负向情感的识别精准度和回召率,然后求两类情感识别的指标宏平均和微平均。指标计算如表 4 和公式(3)－(5)所示:

表 4　正/负向情感判别混淆矩阵

	算法判定正/负向情感	算法判定非正/负向情感
人工判定正/负向情感	TP_p/TP_n	FP_p/FP_n
人工判定非正/负向情感	FN_p/FN_n	TN_p/TN_n

正/负向情感识别的评测指标分别表示为:

$$P_p=\frac{TP_p}{TP_p+FP_p}\quad R_p=\frac{TP_p}{TP_p+FN_p}$$

$$F1_p=\frac{2*P_p*R_p}{P_p+R_p}\tag{3}$$

$$P_n=\frac{TP_n}{TP_n+FP_n}\quad R_n=\frac{TP_n}{TP_n+FN_n}$$

$$F1_n=\frac{2*P_n*R_n}{P_n+R_n}\tag{4}$$

指标的宏平均为正负情感的评测指标的均值,精准度,回召率和 F1 值分别表示为 P_{macro},R_{macro},$F1_{macro}$,微平均指标计算公式如下:

$$P_{micro}=\frac{TP_p+TP_n}{TP_p+TP_n+FP_p+FP_n}$$

$$R_{micro}=\frac{TP_p+TP_n}{TP_p+TP_n+FN_p+FN_n}$$

$$F1_{micro}=\frac{2*P_{micro}*R_{micro}}{P_{micro}+R_{micro}}\tag{5}$$

4.3.2　实验与结果分析

算法实现采用 Python 语言,主要进行了两个实验。

(1)实验一:语料库均衡性对算法的性能影响实验。实验根据评论人对产品的星级评价,将评论分为好评与差评。好评以正向的“特征－情感”表达为主;差评则包含较多的负向“特征－情感”对。预实验发现,语料库中好评与差评的分布对算法性能有较大影响。为从语料质量层面对算法的实际效用进行检验,实验设定了语料均衡参数 K(见公式 6),并将 K 作为优化因子,探测 K 的不同取值对算法产生的作用。语料数量说明见表 5,实验结果见表 6 和表 7。

$$K = \frac{|Pos_review|}{|Neg_review|} \quad Pos_review \subseteq Corpus \quad Neg_review \subseteq Corpus \qquad (6)$$

表 5 正负面语料数量 （单位：条）

均衡参数 K	大众点评餐饮评论语料		京东电子产品语料	
	好评评论量	差评评论量	好评评论量	差评评论量
1∶9	3 007	22 708	13 169	108 919
1∶3	8 500	11 331	36 533	108 919
1∶1	30 726	22 708	101 905	108 919
3∶1	30 726	22 708	101 905	31 875
9∶1	30 726	3 540	101 905	12 057

结果显示，当语料分布均衡时（K = 1∶1），算法表现最好。京东语料上 $F1_{macro}$ = 0.878，$F1_{micro}$ = 0.907；大众点评语料上 $F1_{macro}$ = 0.827，$F1_{micro}$ = 0.832；随着语料分布偏斜，性能指标下降。分析原因，互信息计算基于词共现，情感倾向相同的词项频繁共现；若“特征－情感”对没有在共现窗口内与情感确定的词项建立关系，“特征－情感” 的极性无法判断；“特征－情感” 的情感状态依据邻近情感词极性进行推断，若支持共现率计算的语料不充分，可导致判断失误。实验中，负面语料偏少的情况下（K＝9∶1），支持负面情感计算的共现关系数量不足，负面“特征－情感”的识别精准度和召回率就受到了影响，京东数据集上 P_n = 0.691，R_n = 0.652；大众点评数据集上 P_n = 0.689，R_n = 0.551，综合指标也最低，京东数据 $F1_{macro}$ = 0.777，$F1_{micro}$ = 0.827；大众点评数据 $F1_{macro}$ = 0.687，$F1_{micro}$ = 0.698。可见，语料的均衡性对点互信息算法性能有显著的影响。基于本实验结果，后续实验取 K = 1∶1，即在均衡语料下进行分析检验。

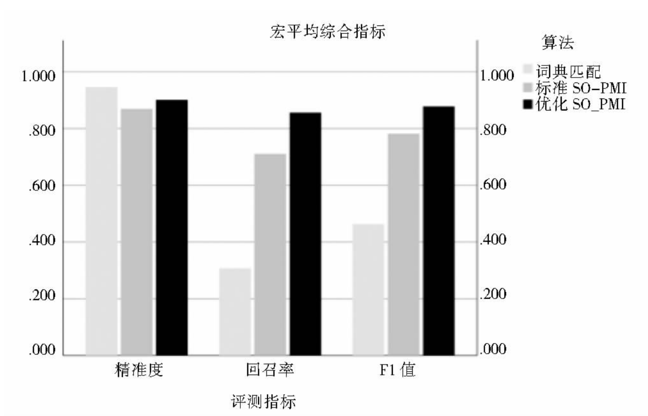
表 6 语料均衡实验结果（大众点评网餐饮评论语料） （单位：条）

K	正向特征－情感		负向特征－情感		宏平均			微平均		
	P_p	R_p	P_n	R_n	P_{macro}	R_{macro}	$F1_{macro}$	P_{micro}	R_{micro}	$F1_{micro}$
1∶9	0.795	0.838	0.791	0.669	0.793	0.753	0.773	0.794	0.765	0.779
1∶3	0.843	0.891	0.897	0.682	0.87	0.786	0.826	0.862	0.8	0.83
1∶1	0.851	0.881	0.882	0.702	0.866	0.791	0.827	0.862	0.803	0.832
3∶1	0.714	0.809	0.702	0.554	0.708	0.682	0.695	0.71	0.699	0.704
9∶1	0.708	0.802	**0.689**	**0.551**	0.699	0.677	**0.687**	0.702	0.694	**0.698**

表 7 语料均衡实验结果（京东商城电子产品评论语料） （单位：条）

K	正向特征－情感		负向特征－情感		宏平均			微平均		
	P_p	R_p	P_n	R_n	P_{macro}	R_{macro}	$F1_{macro}$	P_{micro}	R_{micro}	$F1_{micro}$
1∶9	0.9	0.923	0.781	0.699	0.841	0.811	0.825	0.871	0.862	0.867
1∶3	0.913	0.943	0.835	0.705	0.874	0.824	0.849	0.895	0.879	0.887
1∶1	0.93	0.947	0.872	0.765	0.901	0.856	0.878	0.916	0.898	0.907
3∶1	0.887	0.915	0.752	0.672	0.819	0.794	0.806	0.854	0.849	0.852
9∶1	0.876	0.889	**0.691**	**0.652**	0.784	0.771	**0.777**	0.829	0.825	**0.827**

（2）实验二：基于点互信息的“特征－情感”对的情感极性判断实验。分别在餐饮和数码产品两个领域的均衡语料集上运行改进的 SO_PMI，利用评测语料集对算法识别效果进行检验，参照人工标注，计算精确率、召回率和 $F1$，优化算法性能，同时与词典匹配法及标准的 SO_PMI 算法进行对比。结果如图 2 所示：



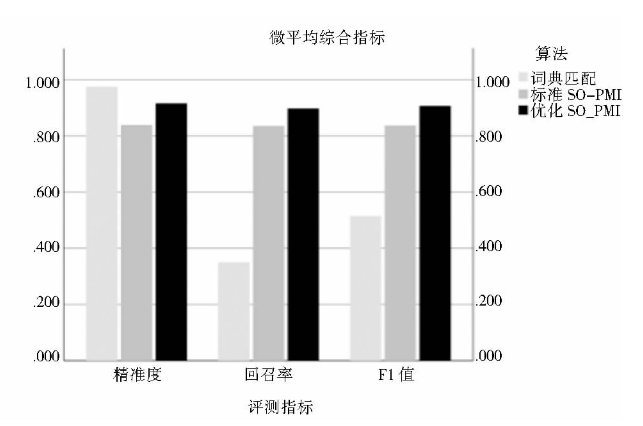


图 2 算法性能比较（京东数码产品评论）

据图 2,在京东评论语料上,改进算法的 $F1_{marco}$ = 0.878;$F1_{micro}$ =0.907,三个算法中改进算法的回召率 R 和 F1 最高。词典匹配法虽然有最高的准确率(P_{macro} =0.946,P_{micro} =0.975),但召回率低(R_{macro} =0.307,R_{micro} =0.350)。显然,词典对未涵盖情感词的极性无法判断。标准的点互信息法较之词典法,准确率略降(P_{macro} =0.869,P_{micro} =0.839),但回召率显著提升(R_{macro} =0.710,R_{micro} =0.835),表明点互信息算法具有较强的发现情感词的能力。而优化了的点互信息算法的准确率和回召率均再次提升,整体表现最好。在餐饮评论语料集上进行实验,改进算法同样表现最佳,如图 3 所示:

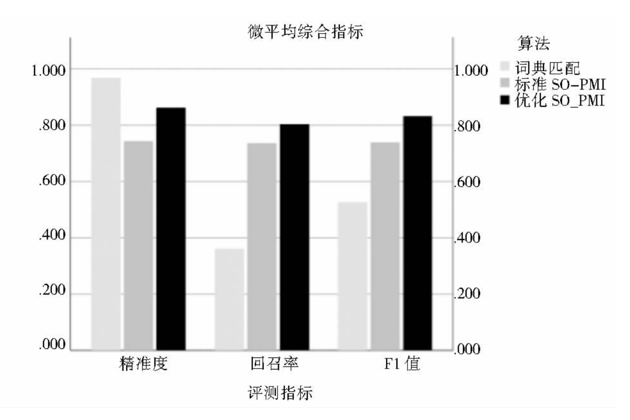


图 3　算法性能比较(大众点评餐饮评论语料)

分析算法提升原因,主要有以下 3 点:

(1)借助大规模均衡语料研判情感表达的极性,突破了词典匹配的局限。优化算法的分析对象定位于“特征－情感”对,但“情感”并不限于情感词项,而是更宽泛的富有情感色彩的表达,如“给力”,“售后都不错”,“现在越来越便宜了”,“都不怎么样”等这些不在词典中出现的情感表达,采用优化算法均可以推断其情感指向,从而能显著提升回召率。

(2)引入情感转折的上下文约束提高了识别精度。优化算法比标准的 SO_PMI 算法有更高的准确率,原因在于标准算法面向情感词,依据词间共现关系推测词项极性,并不考虑特征与情感词间的依存性以及上下文“语境”中的语义转折,结果必定存有偏差,而优化算法借助依存句法分析进行情感语义的转折判断,能够更准确地提炼出待测情感表述与种子词的情感共现关系,提升识别精度。例如“汤味道太咸,但青菜好吃”,采用经典的 SO_PMI,会根据与“好吃”的共现关系推断“味道－太咸”为正向,优化算法则会准确判断出“味道－太咸”的情感极性为负向。

(3)依上下文“语境”准确识别褒贬兼具的情感词极性能提升算法综合效能。自然语言表达中,中性词和褒贬兼具的词项十分常见,易使情感语义产生歧义。如常见的“大,小,高,低”等词项与修饰对象一起方能传达准确的情感指向。本研究提出的算法针对“特征－情感”对展开分析,考虑了“情感”对“特征”的依赖关系,借由大规模语料能够基于上下文约束准确辨识这类情感表达的褒贬含义。以下以“大”为例,列举京东语料下的部分实验结果,见表 8。

表 8　形容词“大”的特征情感对识别结果示例

评论 ID	特征	情感	情感倾向	评论 ID	特征	情感	情感倾向
6064263	内存空间	大	1	6013538	屏幕	大	1
6050609	风扇声音	大	－1	6050603	声音	大	1
6064324	火	大	－1	6064383	手机	大	1
6064327	快递压力	大	－1	6064276	体积	大	1
5907129	铃声	大	1	6033014	效果图	大	1
6064411	面板接缝间隙	大	－1	5907129	字体	大	1
6028738	拍照噪点	大	1	6064308	pro	强大	1
6064399	电脑	非常强大	1	6033023	风扇噪声	有点大	－1
5868634	电脑的质量	都很强大	1	6013511	运行内存	够大	1

据经验,对于数码产品,“大”在评论“内存空间”时为正向;但在形容“风扇声音”时则为负向。“屏幕大”,“电脑的质量强大”为正向评价;“面板接缝间隙大”则判为负向。表8示例表明,算法根据修饰对象给出了适切判断。但结果中,也出现了错判,“拍照噪点大”被判断为正面。分析发现,“拍照噪点”在语料集中只出现一次(“屏幕不够清楚有颗粒感,光线暗一点,拍照噪点大,就是拍照一般般,充电口有点松动,碰一下会有断充问题,京东快递一流”),因出现的语句中“清楚”“一般般”等情感导引词未纳入种子词典,致使“拍照噪点”与正向词“一流”最接近,产生错判。可见,语料均衡性和种子词典的完备性对SO_PMI改进算法的表现作用较大。但总体来看,对近20万均衡评论语料进行分析,*F*1达到近0.9,显著优于词典匹配算法和标准的点互信息算法。

5 结论与后续研究思考

本研究基于大规模领域语料,优化点互信息情感计算方法,以实现特征级情感极性的自动研判。算法依据“特征-情感”对与情感语义明确的种子词的共现关系,同时考虑语句中上下文约束导致的情感转折状态,对待测的“特征-情感”的极性进行推断。真实语料上的实验结果表明,本研究提出的SO_PMI优化算法,性能上优于基于通用词典匹配以及经典SO_PMI情感极性识别法,且算法基于大规模领域语料,不受词典约束,可适用于面向领域的各类型情感分析任务。

总体来看,本文研究贡献可归结为3点:① 将分析对象设定为“特征-情感”对,一定程度上解决词项的“情感歧义”问题。如算法可根据“大”的修饰对象,推断“大”在句中的情感倾向。充分语料的支持下,算法可解决“语境”下的情感分析问题。②算法优化了经典SO_PMI,通过在语句层面提炼词项间共现关系,借助句法依存关联进行情感转折分析,能够准确捕获共现词间的情感一致性,提升SO_PMI的识别精度。③算法为基于语料的情感词极性研判算法。实验揭示出语料充分、正负情感语料分布均衡是算法获得最佳性能的重要原因。且算法基于统计学和自然语言规则,无需过多人工干预,无训练过程,语料获取便利,不受词典约束,这些都使算法具有较强的普适性和迁移性。

需特别说明的是,本研究选取了两个领域进行实验,在两个领域上均获得较理想的实验结果。但比较发现,搜索类产品评论(电子产品)语料集上的表现比体验型产品(餐饮类)语料集上的表现略好。是否是评论语料特点所致?这一问题值得关注。后续研究将拓宽语料范畴,考虑股评、时评、书评等领域,对算法应用做更深入的探讨。

参考文献:

[1] 大连理工大学信息检索研究室.大连理工大学中文情感词汇本体库[EB/OL].[2019-01-10]. http://ir.dlut.edu.cn/EmotionOntologyDownload.

[2] HU M Q, LIU B. Mining and summarizing customer reviews[C]//Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM Press, 2004: 168-177.

[3] LIU B, HU M Q, CHENG J S. Opinion observer: analyzing and comparing opinions on the Web[C]// Proceedings of the 14th international conference on WWW. New York: ACM Press, 2005: 342-351.

[4] POPESCU A M, ETAIONI O. Extracting product features and opinions from reviews[C]//Natural language processing and text mining. London: Springer, 2007: 9-28.

[5] TURNEY P D, LITTMAN M L. Measuring praise and criticism: inference of semantic orientation from association[J]. ACM transactions on information systems, 2003, 21(4): 315-346.

[6] DING X, LIU B, YU P S. A holistic lexicon-based approach to opinion mining [C] // Proceedings of the 2008 international conference on Web search and data mining. New York: ACM Press, 2008: 231-240.

[7] DRACUT E C, YU C, SISTLA P, et al. Construction of a sentimental word dictionary[C]//Proceedings of the 19th ACM international conference on information and knowledge management. New York: ACM Press, 2010: 1761-1764.

[8] WU Y F, WEN M M. Disambiguating dynamic sentiment ambiguous adjectives [C]// Proceedings of the 23rd international conference on computational linguistics. Stroudsburg: Associational for Computational Linguistics, 2010:1191-1199.

[9] 王科, 夏睿. 情感词典自动构建方法综述[J]. 自动化学报, 2016, 42(4): 495-509.

[10] BALAHUR A, MONTOYO A O. Applying opinion mining techniques for the disambiguation of sentiment ambiguous adjectives in SemEval-2 task 18[C]//Proceedings of the 5th international workshop on semantic evaluation. Shroudsburg: Association for Computational Linguistics, 2010: 444-447.

[11] LEK H H, POO D C C. Sentix: an aspect and domain sensitive sentiment lexicon[C]// Proceedings of the 2012 IEEE 24th international conference on tools with artificial intelligence. Washington, DC: IEEE Computer Society, 2012: 261-268.

[12] XIA Y Q, CAMBRIA E, HUSSAIN A, et al. Word polarity disambiguation using Bayesian model and opinion-level features [J]. Cognitive computation, 2014,7(3): 369-380.

[13] WHITELAW C, GARG N, ARAGON S. Using appraisal groups for sentiment analysis [EB/OL]. [2019 - 10 - 01]. http://citeseerx. ist. psu. edu/viewdoc/download? doi = 10. 1. 1. 71. 8147&rep = rep1&type = pdf.

[14] HUANG S L, CHENG W C. Discovering Chinese sentence patterns for feature-based opinion summarization [J]. Electronic commerce research and application, 2015,14(6): 582 - 591.

[15]史伟, 王洪伟, 何绍义. 基于语义的中文在线评论情感分析[J]. 情报学报, 2013, 32(8) : 860 - 867.

[16]聂卉. 隐主题模型下产品评论观点的凝聚与量化[J]. 情报学报, 2017, 36(6): 565 - 573.

[17] ESULI A, SEBASTIAN F. Determining the semantic orientation of terms through gloss classification[C]// Proceedings of the 14th ACM international conference on information and knowledge management. New York: ACM Press, 2005: 617 - 624.

[18] KAMPS J, MARX M, MOKKEN J R, et al. Using WordNet to measure semantic orientations of adjectives [EB/OL]. [2019 - 04 - 11]. http://citeseerx. ist. psu. edu/viewdoc/download? doi = 10. 1. 1. 6. 2534&rep = rep1&type = pdf.

[19] HASSAN A, RADEV D. Identifying text polarity using random walks[C] // Proceedings of the 48th annual meeting of the association for computational linguistics. Stroudsburg: Associational for Computational Linguistics ,2010: 395 - 403.

[20] OLIVEIRA N, CORTEZ P, AREAL N. Stock market sentiment lexicon acquisition using microblogging data and statistical measures [J]. Decision support system, 2016, 85: 62 - 73.

[21] DENG S, SINHA A P, ZHAO H. Adapting sentiment lexicons to domain-specific social media texts [J]. Decision support system, 2017, 94: 65 - 76.

[22] 郗亚辉. 产品评论中领域情感词典的构建[J]. 中文信息学报, 2016, 30(5): 136 - 144.

[23] LABILLE K, GAUCH S, ALFARHOOD S. Creating domain-specific sentiment Lexicons via text mining [EB/OL]. [2019 - 04 - 11]. http://www. sentic. net/wisdom2017labille. pdf.

[24] LI Y, PAN Q, YANG T, et al. Learning word representations for sentiment analysis [J]. Cognitive computation, 2017, 9(6): 843 - 851.

[25] 林江豪, 周咏梅, 阳爱民,等. 基于词向量的领域情感词典构建[J]. 山东大学学报(工学版), 2018, 48(3): 40 - 47.

作者贡献说明:

聂卉:研究问题的提出,研究框架设定,论文完善与修订;

首欢容:数据采集与数据实验,初稿撰写。

Feature-opinion Polarity Identification Based on the Modified PMI Algorithm

Nie Hui　Shou Huanrong
School of Information Management, Sun Yat-Sen University, Guangzhou 510275

Abstract: [**Purpose/significance**] By using corpus-based sentiment analysis, opinion word polarity can be predicted in accordance with its context. The method is significant in applications oriented to specific-domains sentiment analysis tasks since it can improve the prediction accuracy. [**Method/process**] In the paper, context-oriented sentiment polarity identification for emotion expressions was investigated. A Pointwise Mutual Information(PMI) based algorithm was proposed to solve the problem. In terms of PMI, polarity of an emotion expression "feature -opinion" was inferred according to the co-occurrence of the expression with contextual opinion seed words. Furthermore, employing dependence relation analysis to detect sentimental reverse in context; with the modified PMI algorithm, we can predict polarity of emotion expressions in a sentence more accurately. [**Result/conclusion**] The results indicate , compared with the Lexicon-based method and the classical PMI, the modified method performs better. With it, opinion-words unlisted in lexicons can be identified, and context-specific sentimental orientation of an expression can be detected precisely as well. Modifying the macro $F1$ value to 0. 827 and 0. 878 in cater-review corpus and electronic-product review corpus separately. The algorithm, supported by large-scale domain-specific corpus and based on statistics and dependency analysis, is efficient due to convenience for data acquisition, which make it easier be applied in other domain-specific sentimental analysis tasks.

Keywords: sentimental analysis　pointwise mutual information　domain-specific opinion word　context